DOCUMENT RESUME

ED 206 631 TM 810 434

AUTHOR Cotter, Dianne E.; Berk, Ronald A.
TITLE Item Bias in the WISC-R Using Black, White, and Hispanic Learning Disabled Children.
PUB DATE Apr 81
NOTE 50p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Black Students; Elementary Education; Hispanic Americans; *Intelligence Tests; Learning Disabilities; *Racial Bias; *Test Bias; Test Items; White Students
IDENTIFIERS *Wechsler Intelligence Scale for Children (Revised)

ABSTRACT

Item bias in the Wechsler Intelligence Scale for Children (Revised) was investigated using black, white, and Hispanic learning disabled children. Samples of 112 blacks, 126 whites, and 117 Hispanics were matched on subtest raw score and sex. An item-by-group analysis of variance was computed on delta transformed item means for each of nine subtests and two-group comparisons to determine interaction effects. Significant interactions were followed by post hoc comparisons with the Newman-Keuls test. The results indicated bias in four subtests for the black-white comparisons and in three subtests for the Hispanic-white comparisons. Eighteen different items or 22 percent of the items in five different subtests were identified as biased against blacks, whites, or Hispanics. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
[ ] This document has been reproduced as received from the person or organization originating it.
[X] Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED206631

ITEM BIAS IN THE WISC-R

USING BLACK, WHITE, and HISPANIC LEARNING DISABLED CHILDREN

Dianne E. Cotter and Ronald A. Berk

The Johns Hopkins University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. E Cotter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the

American Educational Research Association

Los Angeles, April 1981

TM 810 434

Abstract

Item bias in the Wechsler Intelligence Scale for Children--Revised (1974) was investigated using black, white, and Hispanic learning disabled children. Samples of 112 blacks, 126 whites, and 117 Hispanics were matched on subtest raw score and sex. An item-by-group analysis of variance was computed on delta transformed item means for each of nine subtests and two-group comparison to determine interaction effects. Significant interactions were followed by post hoc comparisons with the Newman-Keuls test. The results indicated bias in four subtests for the black-white comparisons and in three subtests for the Hispanic-white comparisons. Eighteen different items or 22% of the items in five different subtests were identified as biased against blacks, whites, or Hispanics.

There has been increasing concern over the use of individually administered intelligence tests to assess the intellectual ability of culturally different groups. Decisions regarding special class placement are made, in part, on the basis of a child's intellectual functioning on these tests. There have been numerous court cases challenging the legality of assigning black and Hispanic children to educable mentally retarded classes on the basis of the results obtained from intelligence tests (e.g., Larry P., et al. v. Wilson Riles, et al., 1972, 1974; Arreola v. State Board of Education, Note 1; PASE, et al., Joseph P. Hannon, et al., Note 2; Covarrubias v. San Diego Unified School District, Note 3; Guadalupe v. Tempe Elementary School District, Note 4; Larry P., et al. v. Wilson Riles, et al., Note 5; Stewart, et al. v. Phillips, et al., Note 6).

In one class action suit brought on behalf of bilingual Mexican-American children (Diana v. California State Board of Education, Note 7), plaintiffs charged that intelligence tests are "prejudicial to the children in regard to their native language, cultural background, and normative standardization" (Focus on Exceptional Children, 1972, p. 4). Implicit in the Guadalupe case was the assumption that standardized intelligence tests are biased against bilingual children. Plaintiffs in the Stewart, Larry P., and PASE litigation contended that individual intelligence tests are racially and culturally biased against blacks, thereby denying blacks their rights of equal protection under the law in violation of their fourteenth amendment rights.

Legislation at the federal level has had an impact on assessment procedures. For example, the Civil Rights Act of 1964 makes it mandatory that validity data be reported for both blacks and whites (Clark & Plotkin, 1971). The Education for All Handicapped Children Act of 1975 (Public Law 94-142) mandates that assessment instruments be "selected and administered so as not be racially or culturally discriminatory" (p. 781). The Rules and Regulations (1977) for this act requires that the instruments be "validated for the specific purpose for which they are used" (p. 42496). Until recently, validation studies of the Wechsler Intelligence Scale for Children--Revised (WISC-R) have been conducted primarily with normal subjects. However, research carried out on atypical populations (e.g., learning disabled, educable mentally retarded, and emotionally handicapped) is needed in order to comply with the legal mandate of Public Law 94-142.

In view of the social implications of this issue the investigation of bias in intelligence tests is crucial. At present the most widely used individually administered intelligence test is the Wechsler Intelligence Scale for Children--Revised (Wechsler, 1974). Previous bias studies of the WISC-R have focused on its predictive and construct validity across social, racial, or ethnic groups (Dean, 1980; Gutkin & Reynolds, in press a, in press b; Reschly, 1978; Reschly & Sabers, 1979; Reynolds, 1980; Reynolds & Gutkin, in press a; Reynolds & Hartlage, 1979; Sandoval, 1979; Vance & Wallbrown, 1978; Reynolds, Note 8) and across sex (Reynolds, 1980; Reynolds & Gutkin, in press b; Reynolds, Gutkin, Dappen, & Wright, 1979;

Sandoval, 1979; Reynolds, Note 8). This concentration on differential psychometric properties does not furnish evidence on whether specific items within each subtest are biased. Among numerous reasons for studying bias at the item level (see Berk, in press), item bias analysis provides the most basic and direct approach for answering popular charges of cultural bias.

The use of item bias detection methods is especially important in the measurement of content domains that are not well specified (Shepard, Camilli, & Averill, Note 9). That is, when attempts are made to measure psychological constructs, such as intelligence, bias may result because one has no alternative but to conjecture whether or not a particular item is, in fact, measuring a particular construct. Whereas a particular item may function as intended for one group, in that it measures a particular psychological construct, it does not necessarily follow that the same item will measure the same construct (or any construct) in a second group. If inferences regarding the relationship of an item to a particular construct are less valid or less sound for one group as compared with another, bias may result. Therefore, the purpose of this study is to determine whether individual items on the WISC-R function differentially for black, white, and Hispanic learning disabled children.

Method

Subjects

The subjects consisted of 112 black, 126 white, and 117 Hispanic learning disabled children between 10 and 10.33 years of age in a

large metropolitan county in Florida. All of the children met the eligibility requirement for learning disability placement as set forth by the Florida State Board of Education (SBER 6A-6.341). They were concentrated in lower-middle, middle, or upper-middle residential areas. This was determined on the basis of each child's home address and 1979 up-dated census tract data. The index used for the socio-economic classification was the mean family income per tract.

Instrument

The WISC-R is an individually administered intelligence test for use with children between the ages of six and 16 (Wechsler, 1974). It consists of 10 regular and two supplementary subtests. Nine of the 10 regular subtests were analyzed for bias in the present study. The Coding subtest was eliminated because it is a timed copying task that does not lend itself to item analysis.

The total number of points a child can earn on any item depends on the particular subtest. There are both dichotomous and multipoint items. Scaled score equivalents of raw scores are obtained on the basis of age with each particular age level divided into 4 month intervals. For example, at the 10 year old level the first age interval ranges from 10 years, 0 months, 0 days to 10 years, 3 months, 30 days. This is the age interval used in the study.

All of the children were administered the WISC-R by certified school psychologists. The bilingual children were evaluated by bilingual psychologists. (At the present time only an English version of the WISC-R has been published; therefore, the manner in which particular WISC-R items may have been translated from English to

Spanish was at the discretion of each bilingual psychologist.)

Design

One primary concern in item bias studies using the analysis of variance (ANOVA) model is whether or not a significant item-by-group interaction is a reflection of bias or merely a reflection of initial ability differences between the subpopulations being studied. By matching on a relevant cognitive variable such as ability (i.e., subtest raw score), artifactual instances of bias can be reduced and racial-ethnic bias can be examined in a more pure form. Matching on subtest score also retains score variability within the racial-ethnic groups and eliminates variability between the groups.

Although matching on an internal ability variable a priori to the item bias analysis was not introduced into the literature until the late 1970s (see Jensen, 1977; Ironson, Note 10; Plake & Hoover, Note 11), it has since proven useful in satisfying the basic premise involved in the identification of biased items. That is, item bias procedures attempt to identify items that are functioning differently for different racial-ethnic groups of the same ability.

Therefore, in order to add precision to the ANOVA model, a matched-groups design was employed. The children, who were all the same age (within four months), were matched on sex and subtest raw score.[1] This resulted in between 60 and 88 black-white pairs and between 56 and 98 Hispanic-white pairs for the nine different WISC-R subtests. The total number of children and the distribution by sex for the two comparison groups and all subtests are displayed in

Table 1.

Insert Table 1 about here

Judgmental Analysis

Inasmuch as the use of a panel of experts to review individual items for facial bias has been strongly recommended by researchers and test publishers [e.g., Anastasi, 1976; Carlton & Marco (Educational Testing Service), in press;Cole & Nitko, 1981; Green (CTB/McGraw-Hill), in press; Lenke (The Psychological Corporation), in press; Scheuneman, in press; Tittle, in press], two separate committees were formed in order to analyze facial bias in the WISC-R. Six black educators (teachers, administrators) independently reviewed all of the items on the WISC-R (not just those items analyzed in the present study). This review was then followed by a group discussion. A committee of three Hispanic educators (two teachers and one administrator) also discussed all of the items and reached decisions regarding individual item biases.

Statistical Analysis

If one assumes that the trait being measured is normally distributed in the population, then item difficulty can be expressed on an interval scale (Jensen, 1980, p. 439). [Thurstone (1925), through his method of absolute scaling, has demonstrated that intelligence is normally distributed for children at least up until the age of 14.] Unfortunately raw score item means or p values for

dichotomous items do not represent item difficulty on an interval scale. However, particular transformations of those means can produce such a scale. This was accomplished by transforming the original scale first to a z scale, and then to a delta (Δ)scale. The latter has a mean of 13 and a standard deviation of 4 (see Angoff, in press; Angoff & Ford, 1973). For the comparison analyses to detect subtest and item bias in the WISC-R, the delta scale has certain advantages; (1) it permits linear transformations for interval level statistical methods such as analysis of variance; (2) the standard error of Δ is constant for all levels of item difficulty, and this property can greatly simplify some statistical operations (Jensen, 1980, p. 440); (3) delta values are sensitive to ability changes in a given sample (Conrad, 1948); and (4) deltas avoid the false identification of item-by-group interactions for items that are either very easy or very difficult (Merz, 1976; Potthoff, Note 12).

An item-by-group analysis of variance with replications across items was computed for each of the nine subtests and two-group (i.e. black vs. white and Hispanic vs. white) comparisons. A significant interaction effect was interpreted as a flag of possible bias at the subtest level, i.e., item performance was not homogeneous across groups. The power of the F test to reject the null hypothesis for this interaction was estimated to be .39 for the shortest subtest, Object Assembly (4 items), where: $n = 72$, $\alpha = .10$, effect size was expected to be small (ES = .10), and the error rate ratio ($\alpha:\beta$) indicated Type I error to be six times more serious than Type II.

The power value was approximately .87 for the longest subtest, Picture Completion (22 items), where: $n = 98$, $\alpha = .10$, ES = .10, and $\alpha : \beta = 1.3 : 1$. Finally, contingent upon a significant interaction, the significance of bias at the item level[2] was determined by post hoc comparisons of the item difficulties using the relatively powerful Newman-Keuls test.

Committee Review of Statistically Identified Biased Items

Item review following a statistical bias analysis has been suggested by a number of investigators (see, for example, Scheuneman, in press; Fishbein, Note 13; Scheuneman, Note 14) as it may provide clues to the nature of the bias. However, it is not uncommon for committee members not to find logical explanations for the statistical item bias results (Plake, Hoover, & Loyd, 1980; Scheuneman, in press; Ironson, Note 10).

Three committees were formed in an attempt to ascertain whether the item content (or item translation in the Hispanic-white analysis) could explain the differential (racial-ethnic) group responses to those items identified as biased according to the statistics. The first committee consisted of nine black educators (teachers and administrators); the second committee consisted of nine Hispanic educators (four teachers and five psychologists); and the third committee consisted of nine white educators (teachers and administrators). The black and white committees met as a group and discussed each item separately. The four teachers in the Hispanic committee met as a group, whereas the five psychologists responded individually

to each of the statistically identified biased items.

Results

Judgmental Analysis

Black committee review. Five of the six black committee members felt that none of the WISC-R subtests or items were either racially or culturally biased against blacks. The other committee member felt that the Block Design subtest was biased against blacks due to "economic factors." In the words of this teacher, "The greater percentage of black parents will not buy activities of this nature to allow their children to play and practice, thereby lessening success."

Hispanic committee review. The Hispanic committee identified the following eight items, listed by subtest, as being biased against Hispanics:

Information

Item 6: baby cow ("A non-native speaker might not know this. He [or she] would know kitten or puppy.")

Item 16: light bulb ("This item would discriminate against Latin Americans. Latin Americans would be more familiar with Latin American poets, inventors, etc.")

Item 20: ton ("The metric system is used in Latin America.")

Item 24: American man ("Latin Americans are shorter in stature--probably 5'5" on the average--few taller.")

Comprehension

Item 1: cut finger ("A small Spanish child will go to his

[or her] mother.")

Item 3: smoke ("Here again, a small Spanish child will run home and tell his [or her] parents, and they will call the fire department.")

Item 11: inspect meat ("They don't have meat inspectors in Latin America.")

Item 12: charity ("In Latin America you'd see lots of beggars --especially children. There it is better to give to the beggars than to charity, whose directors would probably pocket the donations.")

Statistical Analysis

Black-white comparisons. The analysis of variance results for the black-white comparison group are presented in Table 2. Significant item-by-group interactions were found for the Information ($p < .01$), Similarities ($p < .10$), Comprehension ($p < .01$) and Picture Completion ($p < .10$) subtests. The percentage of subtest variance attributable to each interaction is indicated in Table 3. The percentages for the four significant interactions were all less than 2%. Plots of these disordinal interactions with the item mean differences between the groups[3] are displayed in Figure 1.

Insert Tables 2 & 3 about here

Insert Figure 1 about here

Seven of the 64 items (11%) in the four subtests were identified as biased against blacks using the Newman-Keuls test. Those items are listed by subtest as follows:

Information

Item 6: baby cow ($p < .01$; ES = .41)

Item 12: America ($p < .05$; ES = .35)

Similarities

Item 12: scissors and copper pan ($p < .05$; ES = 1.35)

Comprehension

Item 8: license plates ($p < .05$; ES = .41)

Item 9: criminals ($p < .01$; ES = .65)

Picture Completion

Item 20: screw ($p < .05$; ES = .30)

Item 25: profile ($p < .05$; ES = .53)

Six of the 64 items (9%) were identified as biased against whites. They are listed by subtest as follows:

Information

Item 11: seasons ($p < .01$; ES = .53)

Comprehension

Item 3: smoke ($p < .01$; ES = .54)

Item 6: fight ($p < .05$; ES = .44)

Picture Completion

Item 12: man ($p < .05$; ES = .45)

Item 18: scissors ($p < .05$; ES = .31)

Item 19: girl ($p < .05$; ES = .36)

Hispanic-white comparisons. The analysis of variance results for the Hispanic-white comparison group are shown in Table 2. Significant interactions were found for the Information ($p < .05$), Vocabulary ($p < .01$), and Comprehension ($p < .05$) subtests. The percentage of variance attributable to each subtest interaction is presented in Table 3. The percentages for the three significant interactions were all less than 2%. Plots of these disordinal interactions with the item mean differences between the groups are shown in Figure 2.

Insert Figure 2 about here

Three of the 48 items (6%) on those three subtests were identified as biased against Hispanics. Those items are listed by subtest as follows:

Information

Item 9: bacon ($p < .05$; ES = .38)

Vocabulary

Item 15: prevent ($p < .05$; ES = .50)

Comprehension

Item 16: cotton ($p < .05$; ES = .60)

Four of the 48 items (8%) were identified as biased against whites. They are listed by subtest as follows:

Information

Item 11: seasons ($p < .05$; ES = .49)

Vocabulary

Item 17: nuisance ($p < .05$; ES = .52)

Item 19: hazardous ($p < .01$; ES = .74)

Comprehension

Item 3: smoke ($p < .01$; ES = .81)

Committee Review of Statistically Identified Biased Items

Black committee review. Among the seven items identified statistically as biased against blacks, only Comprehension item 9, criminals, was felt to be so biased by the black committee. The majority of the committee members felt that blacks have had many more negative experiences than whites with the police and the law. The general feeling was that many blacks have been judged to be criminals, not necessarily because of wrong doing, but because of their color. Therefore, a black child who answers this question would be at a disadvantage, compared with a white child.

Among the six items flagged statistically as biased against whites, only Picture Completion item 19, girl, was felt to be biased against whites by the committee. The rationale was that, due to the different hair styles of blacks and whites, a black child would notice the missing ear more than a white child. Comments by committee members such as, "The hair covers the ears of whites" and "Blacks would notice ears more because their ears show more," were generally agreed upon.

Hispanic committee review. The Hispanic committee identified factors that they felt contributed to the bias of two of the three

items that were identified statistically as biased against Hispanics. Information item 9, bacon, was consensually agreed upon as biased against Cubans, Peruvians, and Puerto Ricans due to the cost and scarcity of bacon in these countries. "Cuban children don't even hear of bacon in Cuba. The first time they even hear of bacon is when they get to the United States. Also, most Cuban-Americans rarely, if ever, eat bacon."

The majority of the Hispanic committee members also felt that Comprehension item 16, cotton, was biased against Cubans because "little to no cotton is grown in Cuba." With either limited or no exposure to cotton fields, there may be little discussion generated about cotton, cotton products, and cotton by-products. "Cuban children have not had much experience with cotton or cotton fiber. They would be more familiar with such things as sugar and sugar cane, tobacco, coffee, pineapple, mango, and papaya."

Among the four items that were identified statistically as biased against whites, the Hispanic committee felt that only Vocabulary items 17 and 19 (nuisance and hazardous) were, in fact, so biased. Estorbo, the Spanish translation of nuisance, was judged to be much easier for Hispanic children to define than nuisance for non-Hispanic children. Further, the committee agreed that the word estorbo (and the verb estorbar) is "very common in the Cuban language and Cuban culture." In contrast, hazardous has no Spanish language equivalent. The closest translation is peligroso, which means dangerous. The Hispanic child merely has to give that translation to receive credit.

White committee review. Among the six items identified statistically as biased against whites, only Comprehension item 3, smoke, was felt to be so biased by the white committee. No plausible explanation for this bias was given.

Among the seven items identified statistically as biased against blacks, Comprehension item 8, license plates, and item 9, criminals, were felt to be biased against blacks. Although numerous opinions were offered as to why item 8 could be biased, there was no agreement on any one particular reason. Item 9 was considered to be biased against blacks because, in general, "blacks may perceive the justice system as favoring whites."

Discussion

This section integrates all of the preceding judgmental and statistical evidence in order to explicate why particular subtests and items within those subtests seem to be: (a) biased against blacks, (b) biased against Hispanics, or (c) biased against whites. The discussion of subtest results according to this organization will focus first on the interaction effect that accounted for the largest percentage of variance and end with the effect that accounted for the smallest. Related research on bias will also be examined.

Bias Against Blacks

Comprehension subtest. The interaction effect for the Comprehension subtest accounted for 1.63% of the total subtest variance. Clearly, the item performance of blacks and whites was not homogeneous, i.e., the items did not retain the same rank order of

difficulty, and the same relative differences among the items were not consistent across groups. Sandoval (1979) also obtained a significant interaction for this subtest for his Anglo-American versus black comparison. The amount of explained variance attributable to that interaction was only .14%.[4] Inasmuch as Sandoval did not control for initial ability differences between the racial-ethnic groups, that percentage may reflect ability differences and/or cultural differences.

With regard to the specific items, item 8, license plates, and item 9, criminals, were identified as biased against blacks. The effect sizes for these two items were found to be small and medium, respectively. These results, in general, confirm Sandoval's (1979) findings that items 8 and 9 were significantly more difficult for blacks than for Anglo-Americans; in fact, they were among the items that discriminated most between these groups.

The black committee that reviewed items 8 and 9 for possible sources of bias felt that only the "criminals" item was biased against blacks, whereas the white committee felt that both items were biased. Curiously, although Williams (see PASE, Note 2) has criticized item 4, policemen, as culturally biased, he has not cirticized item 9. Viewed in terms of facial bias, it would seem that both items might elicit similar criticism. However, the results of the judgmental and statistical analyses in this study do not furnish evidence to support the claim of cultural bias in item 4.

Information subtest. The interaction effect for the Information subtest accounted for 1.60% of the total variance. Sandoval (1979)

also found a significant interaction for this subtest with .38% explained variance.

Item 6 on the subtest, baby cow, was flagged statistically as biased against blacks, but the effect size was small. This finding of bias is consistent with the direction of item difficulty differences reported in two other studies (Sandoval, 1979; Sandoval, Zimmerman, & Woo-Sam, Note 15). Furthermore, the magnitude of the differences appeared to be a function of age, with the discrepancy between black and white performance being greater for 7.5 year olds than for 10.5 year olds (see Sandoval et al., Note 15).

The results related to item 12, America, are similar, except that Sandoval (1979) did not find a significant difference in item difficulty. It should also be noted that this item has been criticized as (facially) biased against blacks (Larry P. v. Riles, Note 5).

Similarities subtest. The interation effect for the Similarities subtest accounted for 1.16% of the variance. Sandoval's (1979) significant interaction for this subtest yielded .76% explained variance.

Only item 12, scissors and copper pan, was identified as biased against blacks. The statistical evidence, however, is very persuasive. Not only was there a significant difference between the groups, but the effect size was large (ES = 1.35). This supports Sandoval's (1979) finding that this item was one of the 80 (out of 176) most discriminating items in his study. Unfortunately, the black committee could not provide an explanation for this item bias.

Picture Completion subtest. The interaction effect for the Picture Completion subtest accounted for .84% of the variance. Sandoval (1979) also obtained a significant interaction for this subtest with only .18% explained variance.

Item 20, screw, and item 25, profile, were identified statistically as biased against blacks. The effect sizes were small and medium, respectively. In Sandoval's (1979) comparison, item 20 was significantly more difficult for blacks; item 25 showed no significant difference in difficulty.

Others subtests. The Arithmetic, Vocabulary, Picture Arrangement, and Object Assembly subtests were not flagged as biased against blacks. While the Block Design subtest would also fall into this category based on the statistical analysis, one member of the black committee felt that this subtest was biased, and Sandoval's (1979) research revealed a significant interaction for Block Design with 1.08% explained variance, the largest percentage of any subtest in his Anglo-American versus black analyses. Certainly these judgmental and statistical results cannot be easily discounted; however, further investigation also seems warranted in order to obtain conclusive evidence.

Considering the differences in the linguistic backgrounds of blacks and whites (Johnson & Simons, 1972; Seymour & Seymour, 1979), it is often assumed that tests administered in standard English discriminate against blacks (see Bryen & Bartel, 1973). While this assumption might be upheld for tests that require knowledge of syntax,

morphology, and the like, it is probably untenable for vocabulary tests that require the definition of isolated words, especially where the acceptable response is "any recognized meaning of the word" (Wechsler, 1974, p. 161) regardless of the "elegance of expression" (p. 161). At least the findings in this study that pertain to the Vocabulary subtest lend some support to this position. The non-significant interaction with .60% of the variance indicated that item performance was relatively homogeneous across the groups. Inasmuch as the Vocabulary subtest is individually and verbally administered, these results cannot be generalized to paper-and-pencil, group-administered vocabulary tests that require reading and/or writing skills. For example, an item bias study of the Florida Eighth-Grade Testing Program found 20 vocabulary items on the Reading subtest to be biased against blacks (Fishbein, Note 13).

Bias Against Hispanics

Information subtest. The interaction effect for the Information subtest accounted for 1.63% of the total subtest variance. Sandoval (1979) also obtained a significant interaction for this subtest for his Anglo-American versus Mexican-American comparison. This interaction accounted for .26% of the explained variance.

Only item 9, bacon, was identified statistically as biased against Hispanics, with a small effect size. The Hispanic committee felt that this item was culturally and socioeconomically biased against Cubans, Peruvians, and Puerto Ricans. Sandoval et al. (Note 15) found the "bacon" item to be considerably more difficult

for Mexican-Americans than for Anglo-Americans at both 7.5 and 10.5 years of age. Martin (Note 16), in a Spanish translation, adaptation, and standardization of the WISC-R on randomly selected Hispanics (predominantly Cuban-Americans), also reported that this item was relatively more difficult for Hispanics than for whites (in the WISC-R standardization sample). This evidence strongly suggests that the "bacon" item may be biased against Hispanics due to specific item content.

The judgmental analysis of this subtest by the Hispanic committee revealed that four Information items were facially biased against Hispanics: (1) item 6, baby cow; (2) item 16, light bulb; (3) item 20, ton; and (4) item 24, American man. Unfortunately, items 20 and 24 could not be statistically analyzed because they were too difficult for the 10 year-old learning disabled sample to answer. Items 6 and 16 were analyzed, and they were not found to be biased against Hispanics. However, conflicting evidence has been reported in two previous studies. Item 6 was significantly more difficult for Mexican-Americans (Sandoval, 1979) and Canadians (Beauchamp, Samuels, & Griffore, 1978) than for their Anglo-American counterparts; item 16 was significantly more difficult for Canadians, but no significant difference was observed between Mexican-Americans and Anglo-Americans

Vocabulary subtest. The interaction effect for the Vocabulary subtest accounted for 1.58% of the variance. The percentage of the explained variance accounted for by the significant interaction in Sandoval's (1979) study was larger for this subtest (.36%) than for

any other WISC-R subtest in the Anglo-American versus Mexican-American analyses. In a translation and adaptation of the WISC to Puerto Rican-Spanish (Roca, 1955), the lowest rank order correlation between the WISC standardization sample (white) and the Spanish translation was obtained on the Vocabulary subtest. These consistent findings indicate that item performance is not homogeneous across the Hispanic and white groups.

Item 15, prevent, was the only vocabulary word on this subtest that was identified statistically as biased against Hispanics. Similarly, Martin (Note 16) has demonstrated that based on the Cuban-Spanish version of the WISC-R the rank order of item difficulty for this item changes, compared with the rank order based on the WISC-R standardization sample. Prevent was more difficult for Hispanics than for whites relative to the other items in the subtest. However, despite these findings and the medium effect size of the difficulty difference in the present study, the Hispanic committee was unable to offer a plausible explanation for why prevent was significantly more difficult for Hispanics. Neither the item content nor the Spanish translation, prevenir, appears to contribute to the performance difference.

Comprehension subtest. The interaction effect for the Comprehension subtest accounted for .94% of the variance. Sandoval (1979) also reported a significant interaction with .16% explained variance for his Anglo-American versus Mexican-American comparison.

Only item 16, cotton, was identified statistically as biased

against Hispanics. The Hispanic committee felt that this item was biased because Hispanics frequently come from countried that do not grow cotton, at least not to the extent that it is grown in America. Inasmuch as Hispanics may not have had equal exposure to the concepts presented by this item, the committee judged it to be biased.

In contrast, Martin's (Note 16) study, using the Spanish version of the WISC-R, found that item 16 was slightly easier for Hispanics than for whites (in the WISC-R standardization sample) relative to the other items. These results are somewhat surprising in that both her study and ours employed samples that consisted primarily of Cuban-Americans from the same country in Florida.

The judgmental analysis of this subtest by the Hispanic committee revealed that four items were facially biased against Hispanics: (1) item 1, cut finger; (2) item 3, smoke; (3) item 11, inspect meat; and (4) item 12, charity. While none of these items were found statistically to be biased, Sandoval (1979) did find items 1 and 11 to be significantly more difficult for Mexican-Americans than for Anglo-Americans; however, there were no significant differences between these groups for items 3 and 12.

Other subtests. The Similarities, Arithmetic, Picture Completion, Picture Arrangement, Block Design, and Object Assembly subtests were not flagged as biased against Hispanics. With the exception of the Similarities subtest, these results support Sandoval's (1979) nonsignificant interactions for these subtests.

Bias Against Whites

Information subtest. Item 11, seasons, was significantly more difficult for whites in both the black-white and Hispanic-white analyses, with a medium effect size in both cases. These results are contrary to the findings in three previous studies: (1) Martin (Note 16) found this item to be relatively easier for whites (in the predominantly white standardization sample) than for Hispanics compared with other items in the subtest, (2) Sandoval (1979) found it to be significantly easier for Anglo-Americans than for either blacks or Mexican-Americans, and (3) Sandoval et al. (Note 15) found it to be considerably easier for Anglo-Americans than for blacks, Bermudians, and Mexican-Americans. In addition, the white committee did not feel item 11 was biased against whites. The weight of this evidence brings into question the above statistical results. However, since the earlier studies did not isolate "item bias," it is difficult to reconcile these differences in terms of bias.

Comprehension subtest. Item 3, smoke, was significantly more difficult for whites in both of the analyses, similar to Information item 11. The effect sizes for the black and Hispanic comparisons were medium and large, respectively. In this instance, Sandoval (1979) reported no significant item difficulty difference between either the Anglo-American versus black or Anglo-American versus Mexican-American comparison. The white committee agreed that the item was biased against whites, but the members could not justify that feeling.

Item 6, fight, was biased against whites in the black-white analysis, with a small effect size. Interestingly, this particular item has been criticized by Williams (see, for example, PASE, Note 2) as culturally biased against blacks. In fact, his comments on the CBS television program "The IQ Myth" (Transcript, CBS Reports, Note 17) have stimulated several discussions of this issue in the literature. For the most part, these polemics have been based on conjecture rather than on empirical evidence. Much of the recent evidence suggests that the item: (1) is easier for whites (Gordon, in press; Larry P. v. Riles, Note 5; Sandoval et al., Note 15); (2) is easier for blacks (Miele, 1979); (3) exhibits no significant difference between whites and blacks (Sandoval, 1979); or (4) is slightly easier for whites in the entire WISC-R standardization sample and easier for non-whites in the age range from 8.5 to 10.5 (Doppelt, Note 18). Considering these results in conjunction with those of the present study leads one to conclude that item 6, which has been cited most often as "an example of serious racial bias" (PASE, Note 2, p. 30), does not appear to be biased against blacks.

Vocabulary subtest. Item 17, nuisance, and item 19, hazardous, were biased against whites in the Hispanic-white analysis. The performance differences on both words produced medium effects. The Hispanic committee that reviewed these items for possible sources of bias felt that, due to changes in item difficulty following translation, these items were biased against non-Hispanics. This supports earlier research findings by Martin (Note 16) where, in a

Cuban-Spanish translation of the WISC-R, the rank order of item difficulty changed for items 17 and 19 compared with the rank order for the WISC-R standardization sample. Although nuisance was only slightly more difficult for whites, hazardous was considerably more difficult for whites relative to the other items on the Vocabulary subtest. Furthermore, Fishbein (Note 13) conducted a study with whites and Cuban-Americans and he observed that "all...instances of bias against whites on the White-Cuban comparison were vocabulary items which resembled the Spanish translation and, therefore, gave an unusual advantage to Cuban American examinees" (p. 7). Angoff and Modu (Note 19) have also noted that "it can hardly be expected...that the translated items will have the same meaning and relative difficulty for the second group as they had for the original group before translation" (p. 2). Clearly, item difficulty changes following a translation can result in item bias against the majority population.

Picture Completion subtest. Item 12, man, item 18, scissors, and item 19, girl, were biased against whites based on the black-white analysis. Small effect sizes were found for all three items. The black committee that reviewed the items identified statistically as biased felt that only item 19 was biased against whites. The rationale was that, due to the different hair styles of blacks and whites, a black child would notice the missing ear more often than a white child. Sandoval (1979) found no significant item difficulty difference between blacks and Anglo-Americans on any of these items.

Conclusions

The overall strategy in this investigation of racial-ethnic bias in the WISC-R consisted of three analyses in the following sequence: (1) a judgmental review to detect facial bias; (2) statistical analyses at the subtest and item levels; (3) a posteriori analysis of the statistical results. In essence, the study began with judgment and ended with judgment. While judgmental procedures have been relegated to a secondary position or omitted completely in most studies of bias to date, this study attempted to integrate the information from both judgmental and statistical analyses in order to provide as must insight as possible into the various types of bias that could be identified (e.g., facial bias, differential item performance) and to produce credible results.

The conclusions that can be drawn from all of the analyses are organized into two sections: subtest bias and item bias. Within each section the salient evidence will be presented in terms of statistical significance and practical significance.

WISC-R Subtest Bias

Statistical significance. The item-by-group analyses of variance for the nine WISC-R subtests yielded seven significant interactions ($p < .10$). In the black-white comparisons, interaction effects were found for the Information, Similarities, Comprehension, and Picture Completion subtests; in the Hispanic-white comparisons, the Information, Vocabulary, and Comprehension subtest interactions were significant. These effects corroborate Sandoval's (1979) find-

ings, even though he did not control for ability differences and his samples were "normal" as opposed to learning disabled.

Considering Sandoval's (1979) lack of control for ability, his results on the Arithmetic subtest are particularly noteworthy. It was the only verbal subtest in the Anglo-American versus Mexican-American analysis that did not yield a significant interaction. In our study, it was one of two verbal subtests in the Hispanic-white analysis that produced no interaction effect. In other words, whether ability differences were controlled or not controlled item performance on the Arithmetic subtest was homogeneous across Spanish language groups.

Practical significance. The meaningfulness of the above effects should be evaluated in light of the purpose of the analyses--to flag subtests that might be biased against blacks, whites, and/or Hispanics. Each of the significant interactions accounted for less than 2% of the respective total subtest variance. This suggests that there is very little bias in the particular subtests. However, inasmuch as test biases of any kind are socially and psychometrically undesirable, whatever bias does exist should be eliminated, if possible. Unfortunately, the information provided at the subtest level indicates only that item performance was not homogeneous across groups; specific sources of bias are not pinpointed. The subsequent, exploratory item-by-item analysis and judgmental review were expected to supply this additional evidence.

The judgmental analysis of the WISC-R by the black committee

revealed that the Block Design subtest, not flagged by the ANOVA, appeared to be biased against blacks. The member of the committee who expressed this opinion justified the bias in terms of "economic factors". Interestingly, this apparent facial bias could be considered in conjunction with Sandoval's (1979) significant interaction for the Block Design subtest.

With regard to the Arithmetic subtest, one should be able to assume that diverse cultural groups have equal or near equal exposure to the skills and concepts presented in this (verbally administered) subtest because arithmetic is specifically taught in the schools. (This assumption is probably untenable for the content of many of the other subtests.) If the assumption is valid, it can serve as one possible explanation for why the Arithmetic subtest does not appear to be biased against Spanish language groups. Weighing the results of Sandoval's (1979) study and this study leads to the conclusion that this subtest seems to be the best unbiased measure of verbal intelligence for nonmathematically disabled Hispanic children.

WISC-R Item Bias

Statistical significance. Using the Newman-Keuls test, 18 different items out of 82 (22%) in five different subtests--Information, Similarities, Vocabulary, Comprehension, Picture Completion--were identified as biased ($p < .05$) against blacks, whites, or Hispanics. In the black-white comparisons, 20% of the 64 items on the four subtests previously flagged were biased; 11% were biased against blacks and 9% were biased against whites. In the

Hispanic-white comparisons, 14% of the 48 items in the three subtests cited earlier were biased; 6% were biased against Hispanics and 8% were biased against whites.

The number of biased items in most of the subtests seems to balance out between the racial-ethnic group and the white group. That is, the biases tended to cancel one another when viewed in terms of subtest scores. This observation, however, is a rationalization rather than a solution to the problem. The bias still exists.

Practical Significance. The magnitude of the difference between the group means of each biased item was assessed in terms of effect size and observed Δ discrepancy. For the majority of these items, medium and large effect sizes were found. The differences for all of the items were greater than 1.00; the differences for Comprehension items 3 and 9 exceeded 2.00. According to Conrad (1948), "a difference between Δ's of 1.00 or more represents a difficulty-difference which is both subjectively perceptible and practically significant" (p. 12).

Furthermore, 8 items in the Information and Comprehension subtests in addition to the 18 mentioned above were judged as facially biased against Hispanics. This brings the total number to 26 items or 32% of the items in five WISC-R subtests that should be considered as suspect and probably biased. Given the results of a few previous studies (e.g., Ironson, Note 10), it is not surprising that most of these items are predominantly verbal rather than performance.

Final Remarks

This investigation of the WISC-R can be distinguished from previous studies not only in terms of its findings, but also in terms

of its methodology. Heretofore, a matched-groups design that controls for ability and sex differences coupled with a statistical item bias analysis have not been employed. While further research on bias in the WISC-R should be conducted to verify the results presented here, especially with samples at other age levels and from other geographic locations, the foregoing evidence cannot be dismissed or rationalized away. It merits the serious attention of The Psychological Corporation in the context of subsequent test revisions and of all psychologists who make decisions based on the WISC-R.

Footnotes

[1]Matching on these variables was first attempted across the three racial-ethnic groups (blacks, whites, Hispanics) for each subtest. However, this triadic matching markedly reduced the sample size and, consequently, the statistical power of each ANOVA. Instead, pairwise matching of blacks with whites and Hispanics with whites for each subtest was employed.

[2]Of the 167 items on the WISC-R, not including Coding or the two supplementary subtests, 110 items were analyzed in the black-white comparison group and 113 items were analyzed in the Hispanic-white comparison group.

[3]The procedure used to plot the interaction effect and item mean differences between the racial-ethnic comparison groups was derived from a suggestion by Stanley (1969). The delta transformed mean differences from the average difference for each group are plotted on the Y-axis in Figures 1 and 2 so that the differences are more easily interpretable visually compared to the typical bivariate scatter-plot.

[4]The estimate of explained variance for the interaction effect reported by Sandoval (1979) should not be confused with the estimate of total subtest variance attributable to the interaction effect reported in this study. The former excludes the error variance term, whereas the latter does not. The "total variance" estimate (eta squared) seems to be more widely used in empirical bias studies (see Jensen, 1980) than the "explained variance" estimate.

Reference Notes

1. Arreola v. State Board of Education. Superior Court of Orange County, California, Civil Action No. 160 577, 1968.

2. Parents in Action on Special Education (PASE), et al. v. Joseph P. Hannon, General Superintendent of Schools in Chicago, et al. U.S. District Court for the Northern District of Illinois Eastern Division, No. 74 C 3586, July 7, 1980.

3. Covarrubias v. San Diego Unified School District. Civil Action No. 70-394-S (S.D. Cal., July 31, 1972), concent decree.

4. Guadalupe v. Tempe Elementary School District. District Court for Arizona, Civil Action No. 71-435, Jan., 1972.

5. Larry P., et al. v. Wilson Riles, Superintendent of Public Instruction for State of California, et al. U.S. District Court for the Northern District of California, No. C-71-2270 RFP, October 11, 1979.

6. Stewart, et al. v. Phillips and the Massachusetts Board of Education. U.S. District Court for Massachusetts, Civil Action No. 70-1199F, 1970.

7. Diana v. California State Board of Education. Civil Action No. C-70-37 RFP (N.D. Cal., Jan. 7, 1970; Jan. 18, 1973), concent decree.

8. Reynolds, C. R. Internal validity of the WISC-R as a measure of intelligence for blacks, whites, males, and females. Paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles, April 1981.

9. Shepard, L. A., Camilli, G., & Averill, M. Comparison of six procedures for detecting test item bias using both internal and external ability criteria. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, April 1980.

10. Ironson, G. H. A comparative analysis of several methods of assessing item bias. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978. (ERIC Document Reproduction Service No. ED 163 072)

11. Plake, B. S., & Hoover, H. D. A methodology for identifying biased achievement test items that removes the confounding in an items by groups interaction due to possible group differences in instructional level. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978. (ERIC Document Reproduction Service No. ED 161 930)

12. Potthoff, R. F. Statistical aspects of the problem of biases in psychological tests (Institute of Statistics Mimeo Series No. 497). Chapel Hill, NC: Department of Statistics, University of North Carolina, 1966.

13. Fishbein, R. L. An investigation of the fairness of the items of a test battery. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, D.C., 1975. (ERIC Document Reproduction Service No. Ed 111 837)

14. Scheuneman, J. D. Ethnic group bias in intelligence test items. Paper presented at the annual meeting of the American Educational Research Association, Torono, March 1978. (ERIC Document Reproduction Service No. ED 162 012)

15. Sandoval, S., Zimmerman, I. L., & Woo-Sam, J. M. Cultural differences on WISC-R verbal items. Paper presented at the annual meeting of the American Psychological Association, Montreal, September 1980.

16. Martin, P. C. A Spanish translation, adaptation and standardization of the Wechsler intelligence scale for children--revised. Unpublished doctoral dissertation, The University of Miami, 1977.

17. Transcript. CBS Reports, "The IQ Myth." April 22, 1975, 1-18.

18. Doppelt, J. "Deposition." Reporters' Transcript, Larry P., et al. v. Wilson Riles, et al., U.S. District Court for the Northern District of California, 1977.

19. Angoff, W. H., & Modu, C. C. Equating the scales of the Spanish-language prueba de aptitude academic and the English-language scholastic aptitude test of the College Entrance Examination Board. Princeton, NJ; Educational Testing

Service, 1973, (ERIC Document Reproduction Services No. ED 073 170)

References

Anastasi, A. Psychological Testing (4th ed.). New York: Macmillan, 1976.

Angoff, W. H. Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press, in press.

Angoff, W. H., & Ford, S. F. Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 1973, 10, 95-106.

Beauchamp, D. P., Samuels, D. D., & Griffore, R. J. WISC-R information and digit span scores of American and Canadian children. Applied Psychological Measurement, 1978, 3, 231-236.

Berk, R. A. Introduction. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press, in press.

Byren, D. N., & Bartel, N. R. The disadvantaged as a social and professional phenomenon. Journal of School Psychology, 1973, 11, 387-401.

Carlton, S. T., & Marco, G. L. Bias methods used by test publishers--Educational Testing Service. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins

University Press, in press.

Clark, K., & Plotkin, L. Aptitude test bias. Science, 1971, 174, 1278-1279.

Cole, N. S., & Nitko, A. J. Measuring program effects. In R. A. Berk (Ed.), Educational evaluation methodology: The state of the art. Baltimore: The Johns Hopkins University Press, 1981.

Conrad, H. S. Characteristics and uses of item-analysis data. Psychological Monographs, 1948, 62 (Whole No. 295).

Dean, R. S. Factor structure of the WISC-R with Anglos and Mexican-Americans. Journal of School Psychology, 1980, 18, 234-239.

Diana v. California State Board of Education. U.S. District Court for the Northern District of California. Focus on Exceptional Children, 1972, 4, 1-11.

The Education for All Handicapped Children Act of 1975 (Public Law 94-142). In United States Code: Congressional and Administrative News (Vol. 1). St. Paul, MN: West Publishing Co., November 29, 1975. Pp. 773-796.

Gordon, R. A. Labelling theory, mental retardation, and public policy: Larry P. and other developments since 1974. In W. R. Grove (Ed.), The labelling of deviance: Evaluating a perspective (2nd ed.). New York: Halsted/Sage, in press.

Green, D. R. Bias methods used by test publishers--California Test Bureau/McGraw-Hill Book Co., In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press, in press.

Gutkin, T. B., & Reynolds, C. R. Factorial similarity of the WISC-R for Anglos and Chicanos referred for psychological services. Journal of School Psychology, in press. (a)

Gutkin, T. B., & Reynolds, C. R. Factorial similarity of the WISC-R for blacks and whites in the standardization sample. Journal of Educational Psychology, in press. (b)

Jensen, A. R. An examination of culture bias in the Wonderlic personnel test. Intelligence, 1977, 1, 51-64.

Jensen, A. R. Bias in mental testing. New York: The Free Press, 1980.

Johnson, K. R., & Simons, H. D. Black children and reading: What teachers need to know. Phi Delta Kappan, 1972, 53, 288-290.

Larry P., et al. v. Wilson Riles, Superintendent of Public Instruction for the State of California, et al. 343 F. Supp. 1306, (N.D. Cal., 1972); Aff'd 502 F. 2d 963 (9th Cir. 1974).

Lenke, J. M. Bias methods used by test publishers--The Psychological Corporation. In R. A. Berk (Ed.), Handbook of methods for

detecting test bias. Baltimore: The Johns Hopkins University Press, in press.

Merz, W. R. Test fairness and test bias; A review of procedures. In M. J. Wargo & D. R. Green (Eds.), Achievement testing of disadvantaged and minority students for educational program evaluation. Monterey, CA: CTB/McGraw-Hill, 1976.

Miele, F. Cultural bias in the WISC. Intelligence, 1979, 3, 149-163.

Parents in Action on Special Education (PASE), et al. v. Joseph P. Hannon, General Superintendent of Schools in Chicago, et al. 49 Law Week 2087, 1981.

Plake, B. S., Hoover, H. D., & Loyd, B. H. An investigation of the Iowa Tests of Basic Skills for sex bias: A developmental look. Psychology in the Schools, 1980, 17, 47-52.

Reschly, D. J. WISC-R factor structures among Anglos, Blacks, Chicanos, and Native American Papagos. Journal of Consulting and Clinical Psychology, 1978, 46, 417-422.

Reschly, D. J., & Sabers, D. Analysis of test bias in four groups with the regression definition. Journal of Education Measurement, 1979, 16, 1-9.

Reynolds, C. R. Differential construct validity of Intelligence as popularly measured: correlations of age with raw scores on the WISC-R for blacks, whites, males and females. Intelligence, 1980,

4, 371-379.

Reynolds, C. R., & Gutkin, T. B. A regression analysis of test bias on the WISC-R for Anglos and Chicanos referred to psychological services. Journal of Abnormal Child Psychology, in press. (a)

Reynolds, C. R., & Gutkin, T. B. Stability of the WISC-R factor structure across sex at two age levels. Journal of Clinical Psychology, in press. (b)

Reynolds, C. R., Gutkin, T. B., Dappen, L., & Wright, D. Differential validity of the WISC-R for boys and girls referred for psychological services. Perceptual and Motor Skills, 1979, 48, 868-870.

Reynolds, C. R., & Hartlage, L. C. Comparison of WISC and WISC-R regression lines for academic prediction with black and white referred children. Journal of Consulting and Clinical Psychology, 1979, 47, 589-591.

Roca, P. Problems of adapting intelligence scales from one culture to another. High School Journal, 1955, 38, 124-131.

Rules and Regulations, Implementation of Part B of the Education of the Handicapped Act. Federal Register, Part II, August 23, 1977, 42, 42474-42518.

Sandoval, J. The WISC-R and internal evidence of test bias with minority groups. Journal of Consulting and Clinical Psychology, 1979, 47, 919-927.

Scheuneman, J. D. A posteriori analyses of biased items. In R. A. Berk (Ed.), Handbook of methods for detecting test bias: Baltimore: The Johns Hopkins University Press, in press.

Seymour, H. N., & Seymour, C. M. Ebonics and Public Law 94-142. Journal of Black Studies, 1979, 9, 449-468.

Stanley, J. C. Plotting ANOVA interactions for ease of visual interpretation. Educational and Psychological Measurement, 1969, 29, 793-797.

Thurstone, L. L. A method of scaling psychological and educational tests. Journal of Educational Psychology, 1925, 16, 433-451.

Tittle, C. K. Judgmental bias methods. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press, in press.

Vance, H. B., & Wallbrown, F. H. The factor structure of intelligence for black children: A hierarchical approach. Psychological Record, 1978, 28, 31-39.

Wechsler, D. W. Manual for the Wechsler Intelligence Scale for Children--Revised. New York: The Psychological Corporation, 1974.

Table 1

Total Number and Distribution of Subjects by Sex for the Black-White and Hispanic-White Comparison Groups for Nine WISC-R Subtests

Black-White Comparison Group					Hispanic-White Comparison Group				
WISC-R Subtest	Males (Pairs)	Females (Pairs)	Total Pairs	Total N	WISC-R Subtest	Males (Pairs)	Females (Pairs)	Total Pairs	Total N
Information	50	19	69	138	Information	40	16	56	112
Similarities	45	21	66	132	Similarities	46	24	70	140
Arithmetic	50	21	71	142	Arithmetic	65	26	91	182
Vocabulary	48	12	60	120	Vocabulary	44	23	67	134
Comprehension	52	17	69	138	Comprehension	61	24	85	170
Picture Completion	66	22	88	176	Picture Completion	71	27	98	196
Picture Arrangement	58	12	70	140	Picture Arrangement	52	15	67	134
Block Design	48	17	65	130	Block Design	54	20	74	148
Object Assembly	56	16	72	144	Object Assembly	63	26	89	178

Analysis of Variance Results of the Black-White and Hispanic-White Comparison Groups for Nine WISC-R Subtests

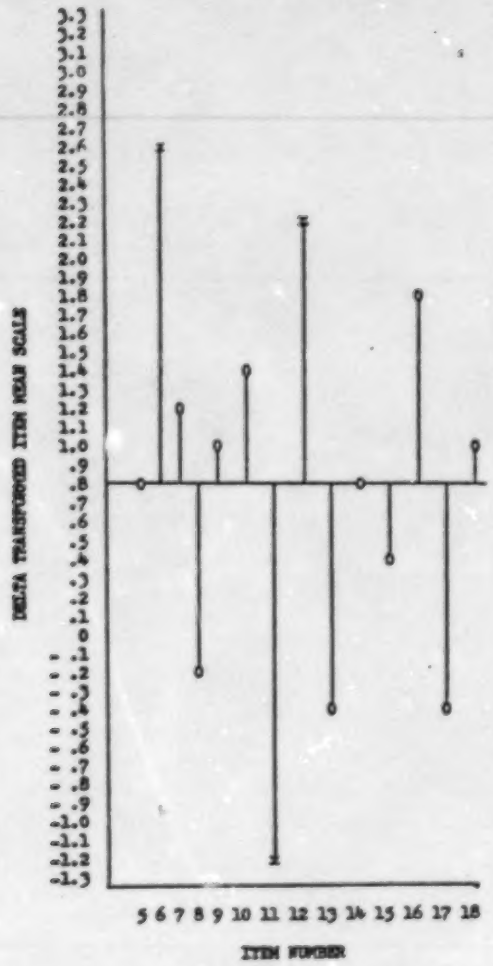
WISC-R Subtest	Source of Variation	Black-White Comparison Group SS	df	MS	F	Hispanic-White Comparison Group SS	df	MS	F
Information	Between Subjects								
	Group (A)	.34	1	.34	.01	1.57	1	1.57	.07
	Error	3499.01	136	25.73		2614.43	110	23.77	
	Within Subjects								
	Item (B)	.00	13	.00	.00	.00	13	.00	.00
	A X B	491.90	13	37.84	2.51***	404.83	13	31.14	2.04**
	Error	26696.78	1768	15.10		21843.16	1430	15.27	
Similarities	Between Subjects								
	Group (A)	1.21	1	1.21	.05	.51	1	.51	.02
	Error	2959.55	130	22.77		3817.13	138	27.66	
	Within Subjects								
	Item (B)	.00	11	.00	.00	.00	11	.00	.00
	A X B	292.59	11	26.60	1.74*	88.23	11	8.02	.53
	Error	21898.62	1430	15.31		22782.13	1518	15.01	
Arithmetic	Between Subjects								
	Group (A)	.10	1	.10	.00	.66	1	.66	.02
	Error	3789.04	140	27.06		6333.86	180	35.19	
	Within Subjects								
	Item (B)	.00	6	.00	.00	.00	6	.00	.00
	A X B	135.96	6	22.66	1.60	25.58	6	4.26	.33
	Error	11866.87	840	14.13		13912.01	1080	12.88	
Vocabulary	Between Subjects								
	Group (A)	2.48	1	2.48	.10	8.25	1	8.25	.28
	Error	3025.72	118	25.64		3941.03	132	29.86	
	Within Subjects								
	Item (B)	.00	16	.00	.00	.00	17	.00	.00
	A X B	192.82	16	12.05	.78	603.59	17	35.51	2.36***
	Error	29147.03	1888	15.44		33751.14	2244	15.04	
Comprehension	Between Subjects								
	Group (A)	2.93	1	2.93	.10	.15	1	.15	.00
	Error	3966.08	136	29.16		7786.30	168	46.35	
	Within Subjects								
	Item (B)	.00	15	.00	.00	.00	15	.00	.00
	A X B	573.17	15	38.21	2.55***	407.43	15	27.16	1.95**
	Error	30529.77	2040	14.97		35070.09	2520	13.92	
Picture Completion	Between Subjects								
	Group (A)	.59	1	.59	.02	2.50	1	2.50	.08
	Error	5641.28	174	32.42		6392.30	194	32.95	
	Within Subjects								
	Item (B)	.00	21	.00	.00	.00	21	.00	.00
	A X B	467.28	21	22.25	1.47*	318.92	21	15.19	1.00
	Error	55491.00	3654	15.19		61926.18	4074	15.20	
Picture Arrangement	Between Subjects								
	Group (A)	.28	1	.28	.01	3.25	1	3.25	.13
	Error	5488.81	138	39.77		3327.86	132	25.21	
	Within Subjects								
	Item (B)	.00	10	.00	.00	.00	10	.00	.00
	A X B	126.73	10	12.67	.93	185.07	10	18.51	1.23
	Error	18848.14	1380	13.66		19891.80	1320	15.07	
Block Design	Between Subjects								
	Group (A)	.25	1	.25	.01	.56	1	.56	.01
	Error	4059.76	128	31.72		7215.36	146	49.42	
	Within Subjects								
	Item (B)	.00	6	.00	.00	.00	8	.00	.00
	A X B	54.80	6	9.13	.68	82.25	8	10.28	.87
	Error	10333.27	768	13.45		13869.78	1168		
Object Assembly	Between Subjects								
	Group (A)	.12	1	.12	.01	.08	1	.08	.00
	Error	3245.51	142	22.86		4489.35	176	25.51	
	Within Subjects								
	Item (B)	.00	3	.00	.00	.00	3	.00	.00
	A X B	59.81	3	19.94	1.45	15.37	3	5.12	.40
	Error	5846.52	426	13.72		6823.20	528	12.92	

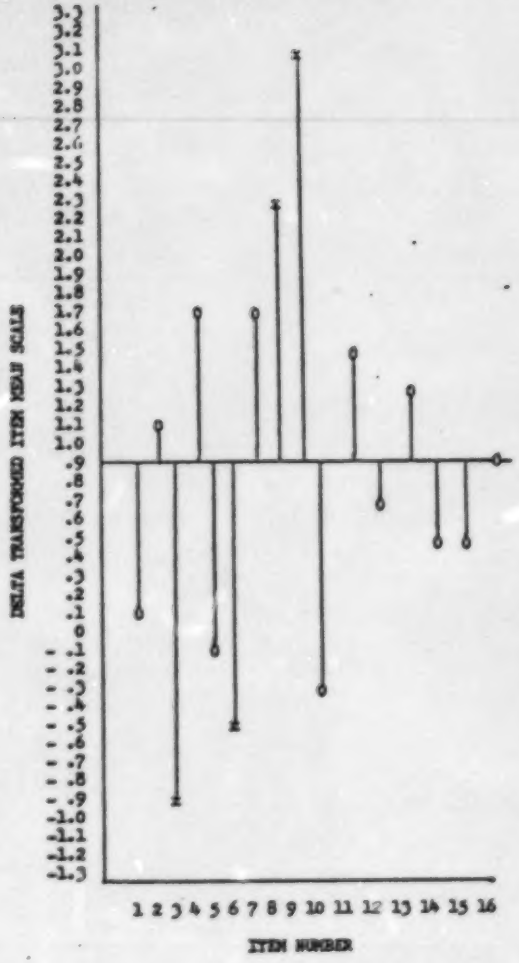
*p < .10. **p < .05. ***p < .01.

Figure Captions

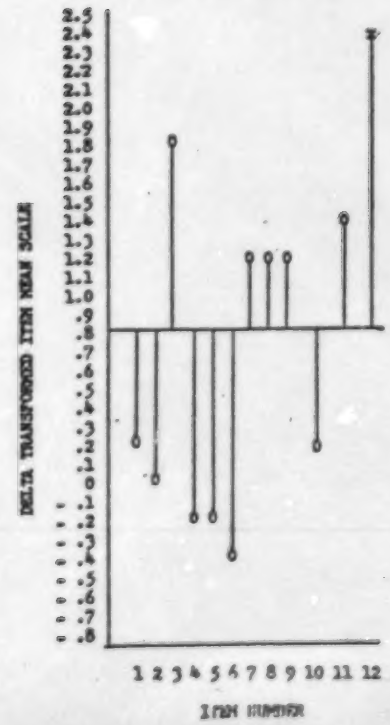
Figure 1. Interaction effects with item mean differences relative to the average-difference line of the black-white comparison group for the Information, Similarities, Comprehension, and Picture Completion subtests. [Points above the horizontal line represent items more difficult for blacks and points below the line represent items more difficult for whites (a point with an "x" denotes a biased item).]

Figure 2. Interaction effects with item mean differences relative to the average-difference line of the Hispanic-white comparison group for the Information, Vocabulary, and Comprehension subtests. [Points above the horizontal line represent items more difficult for Hispanics and points below the line represent items more difficult for whites (a point with an "x" denotes a biased item).]

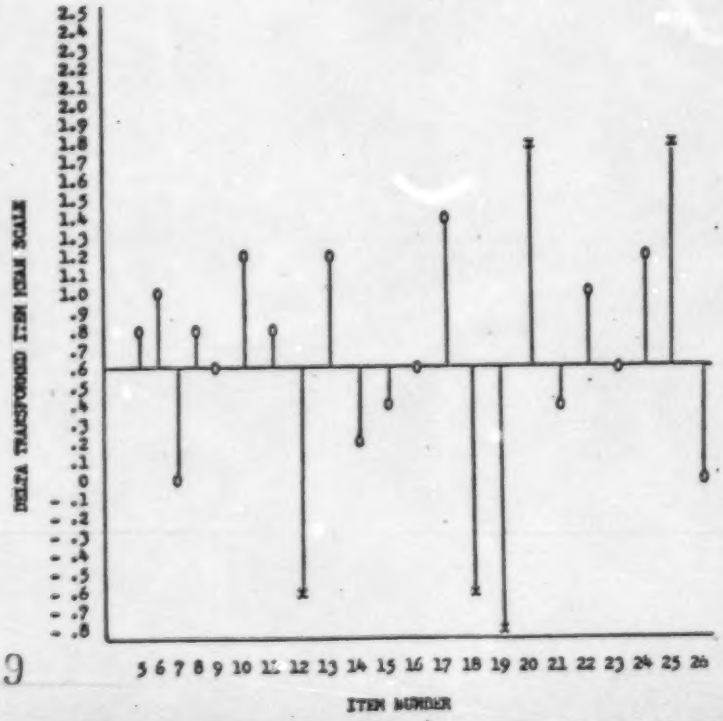




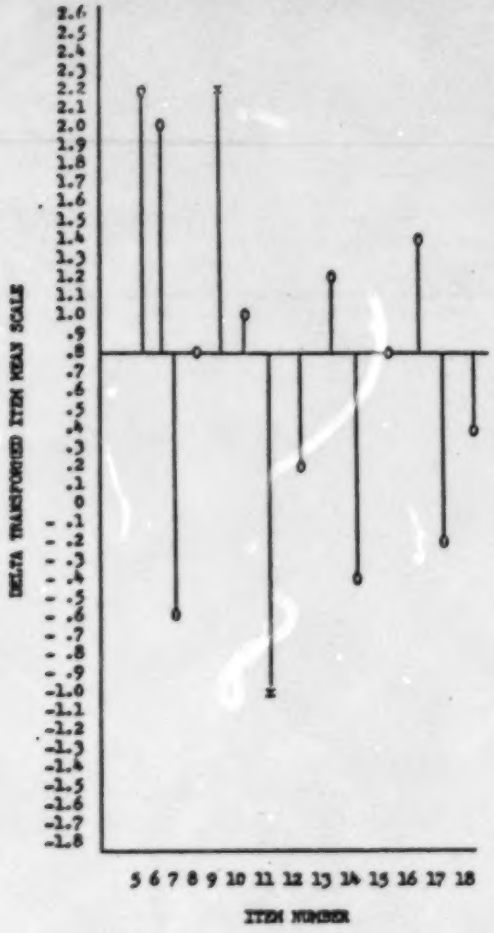
SIMILARITIES



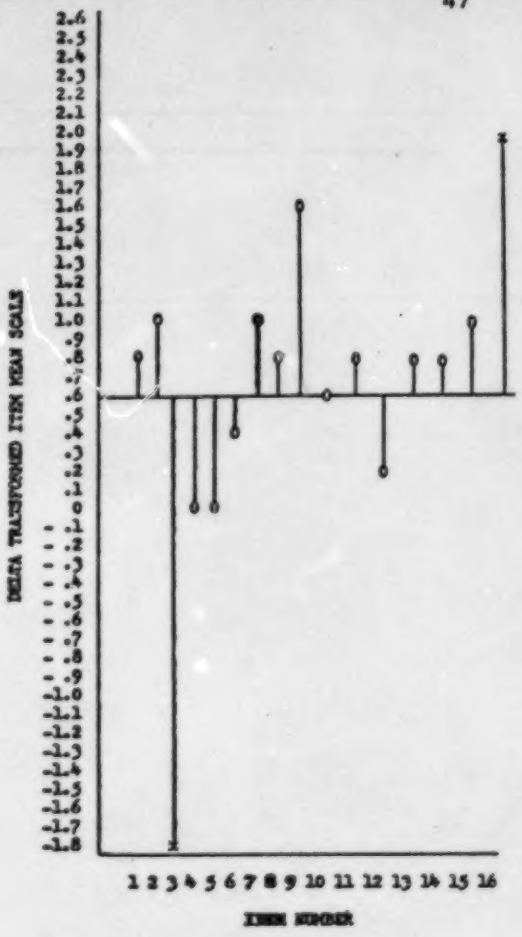
PICTURE COMPLETION



INFORMATION



COMPREHENSION



VOCABULARY